CONTENTS

TWO SHILLINGS AND SIXPENCE NET

ANALYSIS COMPETITION—NINTH " PROBLEM "

The ninth ' problem ' is set by Professor J. N. Findlay of the University of London, and is as follows :

" DOES IT MAKE SENSE TO SUPPOSE THAT *ALL* EVENTS, INCLUDING PERSONAL EXPERIENCES, COULD OCCUR IN REVERSE ? "

Entries (of not more than 600 words), accompanied by stamped, addressed envelope or international stamp voucher if the return of MSS. is desired, should be sent to **The Editor of Analysis by Friday, April 20th, 1956**. *No entries should be sent to Professor Findlay.* Contributors may submit entries either under their own names or a pseudonym. A report, with any winning entries, will be published, if possible, in ANALYSIS for June 1956.

THE EDITOR.

IS PUNISHMENT RETRIBUTIVE?

By K. BAIER

IT would seem that punishment must of its nature be retributive, and also that it cannot be. It must be, for the infliction of hardship on someone is not punishment, unless it is as retribution *for* something he has done. It cannot be, for it makes sense to say that someone was punished for something he did not do. This seemingly simple, but actually quite intricate problem, was recently discussed by Professor A. G. N. Flew in an article entitled ' The Justification of Punishment '[1] and by Mr. A. M. Quinton in a paper entitled ' On Punishment '.[2] Both appear to me to misrepresent the nature of punishment. I shall begin by stating briefly what I hold to be the correct solution, and then point out where exactly they went wrong.

1. To say that someone has punished someone else is to say that someone entitled to administer the penalty for a certain offence has administered this penalty to the person who has been found guilty of this offence by someone with the authority to do so. The question whether or not someone has punished someone else could not even arise unless he belonged to a

group which had the practice of punishing people. We could not say of a group that it had the practice of punishing people unless all the following conditions were satisfied. There must be someone, such as a father or legislator, whose job it is to prescribe or prohibit the doing of certain things or types of thing by certain people, in the form of commands or regulations, someone whose task it is to decree how a person disobeying these commands or regulations shall be treated, someone, such as a father or policeman, entrusted with the task of detecting cases of disobedience, someone, such as a father or judge, charged with meting out the penalty for such disobedience, and someone, such as a father or executioner, charged with administering it. Of course, all these different tasks may be entrusted to one and the same person, as in the case of punishment by a father or teacher.

It should be noticed that ' punishing ' is the name of only a part-activity belonging to a complex procedure involving several stages. Giving orders or laying down laws, affixing penalties to them, ascertaining whether anyone has disobeyed the commands or laws, sentencing persons found guilty, are not themselves punishing or part of the activity of punishing. Yet these activities must be performed and must precede the infliction of hardship if we are to speak of punishment at all. Of course, these activities may take only rudimentary forms. A father does not legislate, but give orders; he does not necessarily affix penalties to breaches of these orders before the breaches occur, but determines the penalty after a breach or disobedience has occurred ; he often does not take much trouble in finding out whether his child really is guilty, nor does he formally " find him guilty " or pronounce sentence. All this is merely tacitly implied, but it is quite definitely implied. It would be just as odd for a father to send his son to bed without supper for being late, if he had found the son not guilty of this—either because the son was not told to be home by a certain time or because he was home by the time mentioned—as it would be for a judge to pronounce sentence on the accused when he has just been acquitted by the jury.

It follows from the nature of this whole " game ", consisting of rule-making, penalisation, finding guilty of a breach of a rule, pronouncing sentence, and finally administering punishment, that the last act cannot be performed unless things have gone according to the rules until then. It is one of the constitutive rules of this whole " game " that the activity called punishing, or administering punishment, cannot be performed if, at a

previous stage of the " game ", the person in question has been found ' not guilty '. The " game " has to proceed differently after the verdict ' not guilty ', from the way it must proceed after the verdict ' guilty '. It is only if the verdict is ' guilty ' that there is any question of a sentence and its being carried out. And if, after the jury has found the accused ' not guilty ', the judge continues as if the jury had found him guilty, then his ' I sentence you to three years ' hard labour ' is not the pronouncement of the sentence, but mere words. If, for some reason, the administration acts on these words, then what they do to the accused is not the infliction of punishment, but something for which (since it never happens) we do not even have a word.

A method of inflicting hardship on someone cannot be called ' punishment ' unless at least the following condition is satisfied. It must be the case that when someone is found " not guilty " it is not permissible to go on to pronounce sentence on him and carry it out. For ' punishment ' is the name of a method, or system, of inflicting hardship, *the aim of which* is to hurt all and only those who are guilty of an offence. For this reason, a system of punishment requires a more or less elaborate apparatus for detecting those who are guilty and for allotting to them the hardship prescribed by the system. To say that it is of the very nature of punishment to be retributive, is to say that a system of inflicting hardship on someone could not be properly called ' punishment ', unless it is the aim of this system to hurt all and only those guilty of an offence. Hence inflicting hardship on a person who has been found ' not guilty ' (logically) cannot be punishing. This is a conceptual point about punishment.

The correct answer to our problem is that punishment is indeed of its very nature retributive, since the very aim of inflicting hardship *as punishment* would be destroyed, if it were inflicted on someone who had been found ' not guilty '. But at the same time, someone may be punished, i.e. have hardship inflicted on him *as punishment*, although he was guilty of no offence, since he may have been *found* guilty without *being* guilty. For all judges and jurymen are fallible and some are corrupt.

2. Flew holds a different view. He says[1] that punishment is of its very nature retributive, but thinks that this applies only to the system of punishment as a whole, not to individual instances of punishing, because " The term ' punishment ' is sufficiently vague to permit us to speak *in single cases and provided these do not become too numerous* (if they do become too numerous,

---

[1] *Loc. cit.*, p. 298.

then *ipso facto* the use, the meaning, of 'punishment' has changed) of punishing a man who has broken no law (or even done no wrong)."[1]

My first point is that Flew has got the analysis wrong. Punishment, whether a system or a single case, is of its nature retributive. Flew says it "would be pedantic to insist *in single cases* that people (logically) cannot be punished for what they have not done",[2] but that "*a system* of inflicting unpleasantness on scapegoats . . . could scarcely be called a system of punishment at all ".[2] These contrasts are misleading. It is not pedantic, but plain wrong, to insist in single cases that people (logically) cannot be punished for what they have not done. But this is not a statistical matter at all. True, a system of inflicting unpleasantness on scapegoats *as such* is not a system of punishment, but then a single case of inflicting unpleasantness on a scapegoat *as such* is not a case of punishment either. In Ruritania, everyone who has been punished during the last year or the last ten years may have been innocent, for in Ruritania the judges and jurymen and the police and prison authorities are very inefficient and very corrupt. A system of punishing people does not turn into a system of inflicting unpleasantness on scapegoats, simply in virtue of the fact that in this system innocent people happen frequently to get punished.

It is surely not true that, if under a certain system of punishment it happens that very many innocent people get punished, the meaning and use of ' punishment ' has *ipso facto* changed. Let us envisage such a deterioration in our own legal system. It is not logically necessary that we should come to know about it. Even now many people claim to have been unjustly condemned. Every now and then we hear of a ghastly judicial error, and there may be many more than we hear of. Think of the many cases in which people accuse the police of having used third degree methods for getting " confessions ". For all I know, a very large percentage of people who are found guilty and are later innocent, are really innocent. At least, this is conceivable. Yet, if it were true, the meaning and use of ' punishment ' could not change *eo ipso*.

Or suppose we knew about it. What would we say ? Simply that judges, police, and so on, were inefficient and/or corrupt, and that very many people got punished wrongly, i.e. unjustly, or by mistake.

Flew may have confused the unsound point he is making

[1] *Ibid.*, p. 299/300.   [2] *Ibid.*, p. 293.

with another point which is sound. Suppose a group had what is properly called a system of punishment. It may then happen once in a while that a judge goes on to " pronounce sentence " even after the jury has found the accused ' not guilty '. Or, to take a more probable case, that a teacher goes on to hit a pupil even after he has realised, perhaps even admitted, that the pupil is innocent. Now, it is true that we would still say that the group had a system of punishment even if such cases occurred, provided they were exceptional. We would not, however, say that they had a system of punishment, if these were not exceptions, but if the group had a system of doing just this sort of thing.

This is true, but again it is not a matter of statistics, not a matter of happening frequently or infrequently. It is a matter of being an exception rather than the rule, in the sense that it is understood to be *a breach of the rule*, rather than merely out of the ordinary. Not merely that judges usually, but not always, discharge the accused when he has been found ' not guilty ', but that it is their *job* or *duty* to do so. If, after the jury has found the accused ' not guilty ', the judge says ' I sentence you to three years' hard labour ', this is not just an unusual case of punishing the man who is innocent, but not a case of punishment at all. And here it would not only not be pedantic, let alone wrong, but perfectly right to say that this case was not a case of punishment.

Flew, I suspect, may have been taken in by the word 'system '. He is, of course, right in saying that a system of inflicting unpleasantness on scapegoats, or a system of ' punishing ' people who had broken no laws cannot be called a system of punishment. This obvious truth, together with the obvious truth that men who have not broken any laws can be and sometimes are correctly said to have been punished, leads him to the view that the solution to the puzzle how one can say the one, but not the other, must be found in the difference between systems and single cases, which he takes to be the difference between the great majority of cases, and single cases.[1]

But it is more complicated than that. The expression " a system *of* ' punishing ' people who had broken no laws " means " system whose *declared and recognised nature* it is to ' punish ' those who had broken no laws ". Hence the importance of ' exception '. If it is still the declared and recognised nature of the group's infliction of hardship on people that it is to be directed to all and only those who are found guilty of an offence, then the cases of inefficient and corrupt judges, and judges

[1] *Loc. cit.*, pp. 302/3.

guilty of flagrant breaches of the law, are clearly exceptions. And while these single cases are exceptions, the whole system can still be called a system *of* punishment, otherwise it is a system of something else.

3. Mr. Quinton, on the other hand, does not think the solution of our problem lies in the distinction between systems and single cases of punishment, but he thinks it lies in the recognition " that ' punish ' is a member of that now familiar class of verbs whose first-person-present use is significantly different from the rest ".[1] As soon as we recognise that while ' I am punishing you for something you have not done ' is as absurd as ' I promise to do something which is not in my power ', we see that ' he punished him for something he had not done ' is no more absurd than is ' he promised to do something that was not in his power '.

My first point is that it is simply not true that ' I am punishing you for something you have not done ' is as absurd as ' I promise you to do something which is not in my power '. It need not be absurd at all. The executioner may whisper it to the man who has been sentenced to death. ' I am punishing you for something you have not done' would be analogous to 'I promise you to do this which is not in my power' only if to *say* 'I am punishing you . . .' *were* to punish you, just as 'I promise you . . .' *is* to promise you. In other words, the verb ' to promise ' is a performatory word, ' to punish ' is not. And if it were used performatorily in ' I hereby punish you . . .' (not, by the way, as Quinton has it, " I am punishing you . . . "), then it would mean the same as ' I hereby sentence you . . . ' and saying it would still not *be* punishing anyone, but merely sentencing him. Thus, Quinton's account is not true of punishing but at best of sentencing.

A similar mistake was made some time ago by Professor H. L. A. Hart in his important paper " The Ascription of Responsibility and Rights "[2] For the view he there expresses[3] is that judicial decisions are ascriptions of responsibility or rights and that these are " performatory utterances ", " non-descriptive statements ", statements not capable of being true or false. But while this is so of some judicial decisions such as pronouncing sentence, it is not true of others such as verdicts. When the jury says ' Guilty ', the accused is " guilty in law " and may have no further recourse against such a judicial decision,

---

[1] *Loc. cit.*, p. 138.
[2] *Proc. Arist. Soc.* 1948–9. Reprinted in Flew, *Logic and Language*, Vol. I.
[3] *Logic and Language.* pp. 155, 157, 159, 161.

but that does not mean that he really is guilty. The verdict
' guilty ' could be a performatory utterance only if uttering
these words were making the accused guilty, as uttering the
words ' I promise ' is making a promise. It might be said that the
jury uses the word in a technical legal sense, different from the
ordinary. When the jury says ' Jury-Guilty ', the accused is not
indeed made guilty, but he is made jury-guilty. But this won't
do, for what does jury-guilty come to? It simply means ' to be
held guilty ', ' to be regarded or treated as guilty '. But this is
not what the jury is asked to decide or what it says. The jury is
asked to give its opinion on whether the accused *is* guilty, not on
whether he is to be *treated as* guilty, for different considerations
might enter into the second question. For the purpose of the
legal consequences, the jury's opinion about his guilt is authori-
tative. Thus, it is not true that the jury says ' Jury-Guilty ', and
thereby makes the accused jury-guilty. What is true is that the
jury says ' guilty ' and thereby makes the accused jury-guilty
that is to be held guilty. Hence the performatory model is out of
place here, for when I say ' I promise ', I am making a promise,
not an ' uttered promise '.

It might be thought that Quinton had seen this point, for
he says[1] " There is an interesting difference here between
' forgive ' and ' punish ' or ' reward ' . . . The three under-
takings denoted by these verbs can be divided into the ut-
terance of a more or less ritual formula and the consequences
authorized by this utterance. With punishment and reward the
consequences are more noticeable than the formula, so they
come to be sufficient occasion for the use of the word *even if the
formula is inapplicable and so improperly used*. But, since the
consequences of forgiveness are negative, the absence of punish-
ment, no such shift occurs ". At first sight, this distinction
between the ritual formula and the consequences authorized by
it might be taken to be the same distinction as the one I have
drawn between sentencing and punishing. But on closer
inspection, this turns out not to be so.

For while to say ' I forgive you ' is indeed to use a formula,
the use of this formula or ritual is not performatory in the way
in which the use of ' I promise ' or ' I hereby sentence you to
. . . ' is. For if I say the latter in the appropriate circumstances
then I have promised or pronounced sentence. But when I say
' I forgive you ' I may *merely say so*. It is moreover wrong to
think that ' I forgive you ' authorizes the non-imposition of
punishment. The assaulted girl, with her last words, may

[1] *Ibid.*, p. 139.

forgive her assailant, but this does not authorize " the absence
of punishment ". Nor is the infliction of punishment authorized
by the formula ' I don't forgive you '. The former indicates the
injured party's intention not to seek revenge, to resume friendly
relations, and so on, the latter the opposite. On the other hand,
the infliction of punishment is authorized by the formula ' guilty '
or equivalent formulae, the non-infliction by the formula ' not
guilty ' and perhaps the Home Secretary's pardon. Thus, the
difference Quinton has in mind is not the difference I have drawn
between pronouncing sentence and punishing.

Lastly, it should be emphasised that it is not true to say that
punishing is the utterance of a ritual formula involving certain
palpable consequences whereas forgiving is merely the utterance
of a ritual formula involving no palpable consequences, and that,
therefore, ' punishing ' sometimes refers merely to the palpable
consequences even when the ritual formula is inappropriate,
whereas this never happens in the case of forgiving. On the
contrary, punishing and forgiving alike are certain kinds of
doings, but they are doings which presuppose the correct
completion of a certain more or less formal procedure culmin-
ating in the finding someone guilty of an offence. If and only if
this procedure has been followed correctly to its very conclusion,
can there be any question of someone's being or not being pun-
ished or forgiven or pardoned. One of the important differences
between forgiving and punishing is that they presuppose different
sorts of formalities. Forgiving is involved only where a man has
been found guilty of an injury. Punishment is involved only
where he has been found guilty of an offence. Many systems of
crime and punishment make injuries offences, but not all offences
are necessarily injuries. Vindictiveness and forgiveness, revenge
and turning the other cheek are individual, punishment and
reward are social ways of dealing with objectionable behaviour.

Although the infliction of hardship on an individual cannot
be called punishment unless it is preceded by his having been
found ' guilty ' of an offence, the procedure leading up to this
finding need only be formally and not materially correct. That
is to say, as long as he was found ' guilty ' in the proper way,
even though he is not in fact guilty, the infliction of hardship
which then follows will be punishment, provided no further
slip occurs.

University of Melbourne

QUANTIFICATION, VARIABLES AND PRONOUNS

By D. W. HAMLYN

IN his article " On What There Is ',[1] Quine maintains that
pronouns are the basic media of reference and that nouns might
better have been named propronouns. His reasons for this
claim are as follows :—Quine maintains that names can be
eliminated from a language by means of an extension of Russell's
theory of descriptions, and this means that we are left with
bound variables, quantifiers and predicative expressions, as, for
example in ' $(x)\phi x$ '. This, he thinks, is desirable in that the
use of names does not of itself guarantee reference. (Elsewhere[2]
he gives reasons for treating the ' ϕ ' in ' $(x)\phi x$ ' as a dummy
predicate and not a variable which is capable of quantification,
and similar remarks apply to ' p ', ' q ', etc. in truth-functional
logic.) Now, logicians have generally sought to indicate that
the ' x ' plays a different role in ' ϕx ' from that which it plays
on its second occurrence in ' $(x)\phi x$ '. Thus in the first case the
' x ' has been called a ' real variable ' and in the second an
' apparent variable '. The assumption behind this is that the ' x '
in ' ϕx ' has a role similar to that in e.g. ' $x+0 = x$ '. On the
other hand it might be said that in these cases the quantifiers
are implicit and it is this view which Quine adopts.[3] The result
of this is the claim that ' ϕx ' by itself does not formalise a
statement, or, in other words, that the ' x ' in ' ϕx ' is not a
dummy-name. ' ϕx ' is what Quine once called a ' matrix ' and
which he now follows Carnap in calling an ' open sentence ',
in that it forms only a part of a statement and is completable by
the addition of quantifiers. In the complete formula (or, as
Quine puts it, closed sentence) ' $(x)\phi x$ ' the ' x ' serves only
as a " mark for cross-reference to a quantifier ".[4] In this, Quine
says, it is like the pronoun ' it '. Thus ' $(x)\phi x$ ' might be inter-
preted as reading ' Whatever x is it is ϕ ',[5] and ' $\exists x.\phi x$ ' as
reading ' There is an x such that it is ϕ.'. The ' it ' in these
interpretations clearly serves only to refer back to the quantifier.
Pronouns such as ' he ' or ' she ' might have a similar use.

However, it is also clear that pronouns like ' it ' are different
from bound variables in one important respect,—to wit, that they
could be used in such a way as to be independent of any expres-

[1] *From a Logical Point of View*, p. 13.
[2] Op. cit., p. 110.
[3] Op. cit., p. 110.
[4] *Methods of Logic*, p. 89.
[5] Cf. *Mathematical Logic* (1947 edition), § 12.

sion corresponding to a quantifier. Consider the following schematic conversation,—" It is very funny." "What is?" " Oh, the film which is on at the local cinema." Now it is of course true that a statement such as the first is generally taken to presuppose some background which enables the ' it ' to be understood. The speaker may have this in mind, but if he does not make it explicit he is liable to be subject to challenge. Nevertheless I do not think that it is true that the statement " It is very funny " is equivalent to " There is something such that it is funny ". The ' it ' is being used to establish reference in the first statement, not to state that there is such ; in this respect it is in no way different from other referring expressions such as those used as names. But I think that it is the fact that they can be used in this way which makes Quine's suggestion that pronouns are the basic media of reference at all plausible although, of course, false. In the quantified formula the variable has no sense apart from the quantifiers, as Quine himself recognises ; but this makes it equivalent to a pronoun only in the sense that pronouns like it can be used to provide cross-references to expressions equivalent to quantifiers. The variable does not provide reference to any entity, but this is provided by the quantifier, or the whole expression of quantification, if it is provided at all. The last clause is important, for the formula ' $(x)\phi x$ ' might, of course, be a valid formula, even if there were no values of the variable ' x '. This could be brought out by rendering ' $(x)\phi x$ ' as ' If anything is an x it is ϕ '. It is worth noting that in this case the ' x ' in the quantifier acts as a dummy class-expression, as it does in the interpretation of the existential quantifier, ' There is an x such that . . .'. But Quine's interpretation of the universal quantifier, in his ' Mathematical Logic ',—' Whatever x is . . .'—would in normal usage be taken as establishing reference ; it would, in its use, presuppose a range of entities, however small, to which ' x ' can apply. It is like saying, " Whoever Jones is, he is important ', and this means that the ' x ' in ' Whatever x is . . .' is to be taken as a dummy-name. The interpretation which Quine offers in his ' Method of Logic ',—' Each thing x is such that . . .' and ' There is something x such that . . .' is ambiguous as regards the status of the ' x '. Nevertheless it is clear that here too it is the clause corresponding to the quantifier that provides the reference. On the other hand, the suggestion that he sometimes makes, that the whole context of quantification, ' $(x) . . . x$ ' is to be interpreted as ' everything ' means that in this case we retain the reference (for ' everything ' would in normal usage be

taken as a referring expression), but give up the claim that the second ' x ' is like a pronoun, in Quine's sense, in which he means third-person personal pronoun. In general, then, it is the clause corresponding to the quantifier which provides reference, where it does so at all, and it is clear that the existential quantifier provides it by asserting it.

If this is so, the variables in quantified formulae do, of course, function like ' it ', but only insofar as they serve to provide cross-references to the quantifiers, and *in no other respect*. Thus the thesis that pronouns are the basic media of reference, if it can be maintained at all, must be maintained on other grounds than those which Quine brings forward. I do not believe, however, that it can be validly maintained at all. It seems to me that *no* type of expression is the basic or special medium of reference, although certain types of expression, e.g. some nouns and pronouns, are typical media of reference in ordinary language. That is to say that it is typically, although not necessarily, these grammatical forms of expression which we use in order to refer to entities. (It is a different, but very interesting question, whether such grammatical forms of expression could be distinguished at all, unless some attention were paid to their typical uses.)

Quine says, " To be assumed as an entity is, purely and simply, to be reckoned as the value of a variable '.[1] Seeing that by ' variable ' he means ' bound variable ', this is equivalent to saying that if something is to be treated as an entity, it must be capable of answering to the description of something, the existence of which is asserted. There is nothing very startling about this, although one might quibble over the details. When he goes on to say, " In terms of the categories of traditional grammar, this amounts roughly to saying that to be is to be in the range of reference of a pronoun ",[1] this, by being less technical, sounds much more startling. But, for reasons that I have given, I think that it is wrong. I also think, however, that it is only a symptom of a more deeply-seated disease, in that it is based upon the assumption that it is possible to make direct translations from ordinary language into logical formulae and vice versa, when the truth is that the formation rules of a logical calculus do not necessarily correspond in any exact[2] way to the grammatical rules of a living language. One indication of this is that within quantificational logic it is permissible to make moves

[1] Op. cit., p. 13.
[2] There must, of course, be *some* correspondence, for the calculus to be called ' logical ' at all. But the logical calculus gets beyond ordinary arguments and reasonings in the way in which mathematics gets beyond the ordinary uses of numbers in e.g. counting.

involving unquantified formulae, which are, on Quine's view, not statements at all, whereas, in ordinary inferences this would by no means be possible. The assumption that such direct translations are always possible cannot but produce metaphysics where no metaphysics is required.

University of London.

CAN THERE BE ARTIFICIAL MINDS?

By JONATHAN COHEN

I

THE first and commonest move[1] in recent controversy about this question has been to compare the familiar mental activities and experiences of men, such as thinking up a joke or imagining an invention, with the actual or possible performances of servomechanisms, computers, and other programmed artifacts. Neurologists, businessmen and generals are all interested in what such artifacts can be designed to do and how they can do it. It seems plausible to suppose that it is this interest which is crystallised in the question ' Can there be artificial minds? ' And that question is then one to be discussed by experts in cybernetics, not by lay philosophers.

The second move[2]—an advance—is to argue that these technical discussions unjustifiably assume that the question is a proper one. Of course, it makes sense to ask whether a robot could duplicate all forms of human behaviour and sensitivity, and this is a legitimate problem for cyberneticists to discuss. But to have a mind of one's own entails a capacity for consciousness, only living things can have conscious and unconscious states, and ' machines do not even belong to the *category* of things which can be dead or alive '. Hence we have a paradox to puzzle us. ' On the one hand it does not seem that there is anything in the construction, constituents or behaviour of the human being which it is essentially impossible for science to duplicate or synthesise. On the other hand there seems to be some important and meaningful descriptions of human behaviour which can never be properly applied to machines.'[3] And we resolve the

---

[1] E.g. A. M. Turing in *Mind*, Oct. 1950 ; L. Pinsky in *Mind*, July 1951 ; and J. O. Wisdom, R. J. Spilsbury, and D. M. Mackay in *Proc. Ar. Soc.*, Supp. Vol. xxvi (1952).
[2] E.g. M. Scriven in *Mind*, April 1953. [3] M. Scriven, ibid., pp. 233-4.

paradox by coming to understand both these apparently con-
flicting truths well enough to recognise that they do not really
conflict.

But I do not think it satisfactory to leave the matter at this
stage. If anyone does use the description ' conscious machine '
or ' intelligent computer ', we are told, one of these terms must
have changed in meaning. So long as we recognise the change,
some philosophers assure us, we shall not be puzzled. And,
presumably, so long as we are not puzzled everything is all right.
But I believe that there are other worth while things which
philosophical analysis can help to achieve or maintain besides
the avoidance of puzzlement. And I want to show how a preva-
lent threat to one of these things seems to be aided if the metaphor
of ' artificial minds ' is allowed to settle into lay speech and be-
come familiar outside the laboratory. In order to bring out how
this comes about I shall concentrate on showing just how far-
fetched this metaphor is. For I shall argue that the knowledge
which we can in principle acquire about any programmed artifact
whatever is of such a type that if we ever acquired knowledge
of this type about a human being we should describe him as
having *no* mind of his own. Indeed, where category-assignments
are at stake, it is often more cogent to argue from differences
and similarities in the types of knowledge attainable than from
the absurdity or good sense of ascribing certain qualities, like
consciousness, in the relevant cases. For when a major category-
shift is being systematically proposed or carried through by any
group of human beings all their relevant notions of what makes
sense and what does not may alter accordingly.

II

Consider, firstly, two comparatively simple situations in
which a cyberneticist might find himself. He has a servomecha-
nism, or a computing machine, with no randomising element,
and he also has a wife. On the one hand, say, he wishes to predict
how his machine is going to behave during the afternoon ;
and, on the other, he wants to predict how his wife will behave.
Knowing the artifact's programme and assuming no breakdowns
he can calculate in principle exactly how it will behave, within
the limits of his true beliefs about what environmental events
will occur to affect it. So far as he knows what information will
be acquired by the feed-back or what questions will be put to
the computer he can in principle deduce the machine's behaviour
from its programme. With regard to his wife, however, he

must make do as a starting point with whatever he has learnt about her character from his experience of her and of other human beings and from the wisdom of novelists, psychologists and friends. From these beliefs, assuming she remains in normal health, he can try to work out how she will behave, within the limits of his knowledge about what environmental events will occur to affect her.

Thus in both cases the logical form of his reasoning is the same. Roughly, both predictions are argued from a set of general propositions coupled with a set of singular propositions which satisfy the antecedents of these hypotheticals. Moreover, in both cases the evidence for the singular propositions is similar in type. For both these sets of propositions are predictions about environmental happenings in the afternoon. But the other two sets of premisses differ from one another in logical status. The cyberneticist's beliefs about his wife's character are not on the same level as his beliefs about his artifact's programme. Only if he found out that he had remembered or recorded this pro-grame incorrectly would he revise his beliefs about its content. But many other events, besides the discovery of a lapse in memory or a slip of the pen, might suggest a revision of his beliefs about his wife's character. Those beliefs were based on past experience and further experience may justify altering them. It is no objec-tion that the cyberneticist may sometimes be startled by what his machine does, i.e. by the logical consequences of its pro-grammes. His surprise is then due to the inadequacy of his prior calculations, whereas there may often be times when no amount of prior reasoning would prevent him from being surprised at his wife's behaviour. Hence there is an asymmetry between the type of knowledge the cyberneticist has about his machine's behaviour and the type of knowledge he has about his wife's.

Let us see if this asymmetry could be remedied by altering details of one or both of the two situations. We might begin by supposing that the artifact contains a randomising element, so that certain features of its behaviour, likely to be activated during the afternoon, are determined by a dice-throw or similar device. The cyberneticist's predictions about his machine would now have to specify certain ranges of alternative events, about which he could say no more than that in each range one and one only of the specified events (or groups of events) would occur. His knowledge of, say, the mechanics of dice-throwing would not be enough for him to predict which event (or group of events) was more probable than the others. But these new predictions as a whole would still have just as much logical certainty as before.

It might well be that the cyberneticist also believed that in certain likely circumstances his wife might do one or other of several alternative actions, while his knowledge of her character was insufficient for him to be able to decide which of these alternatives was the most probable. His predictions about her behaviour would then differ in type from his predictions about his artifact in a way precisely analogous to that in which they differed before.

Someone might now suggest :—' Suppose the cyberneticist built his machine so that once it had been programmed in any way he couldn't " take the lid off " without altering the programme. Suppose, too, that he lost both his records and his memory after programming it. He would then be able to reconstruct its programme only by observing its behaviour, and any predictions he made about the latter would be of the same logical type as predictions about his wife's behaviour.' Well, let us suppose this. It would certainly be hard luck for the cyberneticist. It would cause him a lot of extra work—work of a sort to which he might be unused in his laboratory. But conceivably his memory might return to him or his records might be found by the charwoman, whereas nothing could ever happen which would make his beliefs about his wife's mental character as justifiably certain as his beliefs about the artifact's programme would then be.

' Suppose, on the other hand, that physiologists had discovered a method of " taking the lid off " his wife's brain, nervous system, glands, etc., which enabled surgeons to record every detail of their structure and content. Suppose also that psychologists had found reliable correlations between such features and the character-traits of those people who exhibited them. (After all, you claim to be concerned with everything that is logically possible, not merely with what is biologically possible.) The cyberneticist's knowledge of his wife's character would now depend on these psychological correlations, which might conceivably be just as reliable as the generalisations in electronic theory on which he bases his assumptions about the possibility of programming any machine successfully (i.e. without breakdowns). What now would be the difference in logical type between his two predictions, even if he recovered his memory, or his record, of the artifact's programme? ' Well, it would be the same as the difference between the kind of prediction that was based on direct acquaintance with the programme fed into the artifact and the kind that relied on inferring this programme from observation of the machine's dials, circuits, etc.

' But suppose the surgeons could operate on his wife's brain

in such a way as to alter her character in any desired respect. Surely the cyberneticist could now programme his wife—leaving no contingency to chance whims and making her in all mental respects his ideal woman—just as much as he could programme his artifact?' Certainly : his two predictions about the afternoon would then be altogether of the same logical type. But we should then say, I think, that his wife had no mind of her own, that her husband had already explicitly or implicitly done all her thinking for her, and that the notion of programming an ideal mind was therefore self-contradictory. For consider some of the situations in which we do say that a human being either has, or acts as if he has, no mind of his own.

One situation occurs when we are watching the parade-ground drill of a Guards battalion. ' By force of training ', we might say, ' they have come to let all their thinking be done for them by their officers and n.c.o.'s, and so they move with clockwork precision like mindless machines.' Of course, we should admit that the soldiers are not really without minds of their own, because when they are off duty we know no programme of rules and commands which effectively controls all their behaviour. Again, there might be a monastery in which every facet of day-to-day life was governed by the rule of a religious order. Admirers might say of such monks, ' They have surrendered their own minds in fulfilment of a Higher Purpose ', though no-one would doubt that the monks really had minds of their own. For it would always make sense to talk about their leaving the order, even if this never in fact happened. But I do not see how our cyberneticist's unfortunate wife could ever recover a mind of her own. Her husband's difficulty—suppose he wanted such a recovery—would be logical, not technical. Whatever randomising elements he introduced into her brain, or however much he asked other cyberneticists to programme her and not to tell him what they had done, neighbours who were ' in the know ' would still nod their heads at one another as she walked past and murmur, ' Poor girl, she can never again think for herself or have any emotions that are really her own '.

III

What I have been trying to point out is an opposition between the still familiar concept of mentality and the concept of total subservience to known or knowable rules. The metaphor of ' artificial minds ' clearly destroys this opposition by ascribing mentality to programmed artifacts. That is why such a metaphor

is—outside the laboratory—not merely eccentric and novel, but also undesirable. If we blur our concept of mind in this respect, then, however completely human beings were regimented in thought and action by their government's, party's, or church's ideological programme, it would be false to say that they behaved as if they had no minds of their own. They too could still be said to think for themselves. And that would be a shift in the notion of having a mind of one's own and thinking for oneself of which George Orwell's Newspeak could have been proud.

University of St. Andrews.

PRESUPPOSITIONS

By A. PHILLIPS GRIFFITHS

I

THAT which is presupposed may be the grounds for a belief, or something which enables us to regard certain facts as grounds for a belief. This may lead us to look at presuppositions on the model of suppressed premisses in deductive arguments which must be brought to light and questioned.

A professor may not, in general, doubt his colleagues' honesty. This would be shown if he said with confidence that his monogrammed fountain pen would be returned to him when he left it in the S.C.R. In this case, one would not want to say the professor is presupposing that his colleagues are honest, since he probably knows they are.

But what kind of thing does he know? If one of his colleagues were caught stealing his pen, he would not necessarily have to say he had been wrong all along. He would not have to say this if it could be shown that there were special circumstances involved (for example if the thief had been undergoing some unusually severe mental strain). It would be no reflection on his faith in his colleagues if he admitted that extraordinary circumstances might arise in which one of his colleagues could plausibly be expected to act dishonestly.

In order, therefore, that grounds for a belief that his pen will be returned may be set out which constitute a valid deductive proof, more than " My colleagues are honest " is required to be added. We must also add that those special circumstances

in which a dishonest act would be a plausible possibility do not in fact obtain.

If a presupposition is a suppressed premiss which has not been established, " No special circumstances obtain " must be being presupposed by the professor. For this is a premiss which one would find it very difficult indeed, if not impossible, to establish. Not only is there the difficulty that the list of known possible special circumstances may be impracticably long, there is also the fact that one may never be sure that the known special circumstances are the only possible ones. New causes of dishonesty remain to be discovered, and uniquely new circumstances may arise.

Yet it would seem incorrect to say that the professor is presupposing either that any given special circumstances have not arisen, or that no special circumstances have arisen.

If he were asked "Are you sure that one of your colleagues is not suffering from a mental disease? " he might properly answer "Which one? " or " Why do you say that? " If the answer to these questions is unsatisfactory (e.g. " Well, it's logically possible "), he could say " You should not cast aspersions on my colleagues ". To ask him this question would be to imply a doubt of his colleagues' sanity, and it would be unreasonable to do this unless there were reasons for doubting their sanity.

Where one is not prepared to accept a doubt, one ought not to admit to a presupposition. This case shows that although something may be a necessary addition to a body of evidence if it is to become a valid deductive argument, it does not follow that it is being presupposed, even if it has not been established.

I am not of course suggesting that in general unless one has reasons for doubting something one must accept it as certain. My observations are confined to the particular case. While in general we require reasons for a belief before we can accept it, we require reasons to doubt that a situation is normal. I shall try to explain why.

When the grounds for a belief are held to be good only if some (non-enumerative) generalisation is accepted, the conversion of the grounds plus this generalisation into a deductive argument requires the addition of some such phrase as " in normal circumstances " or " other things being equal ", and a statement that this condition is fulfilled.

" In normal circumstances " differs from " other things being equal ". " Other things being equal, University posts are awarded to candidates with the best manners " does not imply

that University posts are in normal circumstances given to candidates with the best manners. " Other things being equal " warns us that, possibly in every situation, other special circumstances may have to be taken into account ; " in normal circumstances " suggests that normally one need not worry about special circumstances.

" The unusual ", " the abnormal ", " the extraordinary " and " the unexpected " are expressions used to refer to that which we require special circumstances to bring to our notice. It is therefore improper to bring up the question whether a situation is extraordinary unless some special circumstances have given us reason to do so.

II

" The normal ", " the unusual ", and " the extraordinary " do not pick out fixed categories of circumstances. What may be normal in one context may not be so in another ; and this is true of the context of purpose as well as the geographical context.

Our ideas of the normal may suffer change, as circumstances change. But there is also normality with regard to vicissitudes, and this allows us to speak of situations in which we may normally expect circumstances to have remained normal. It would also allow us to construct a series of cases in which it becomes more and more plausible to insist that we should question our concept of the normal.

There will also be a further range of cases for which it is impossible to imagine our idea of the normal suffering change. There are cases where the language we use to pick out the case loses its point if the idea of the normal does not remain the same. These are the philosophically interesting and important cases.[1]

The following set of cases illustrate the above remarks :—

(i) " It will cost you fourpence to go from Marble Arch to Piccadilly " may depend on a published fare list as that which is normal (so that an abnormal situation would be where one is overcharged, for example). Rates of fares may be very changeable, and it may well be proper to say " You are presupposing that the fares haven't changed this week ". The propriety of referring to a presupposition here could become less and less obvious as it becomes normal for fares to remain unchanged. If in future London 'bus fares become iron traditions, there will be no such propriety.

[1] Cf. Wittgenstein, *Philosophical Investigations*, I. 142 and IIv.

(ii) " Smith has burned himself " may be asserted when all that is known is that Smith has put his hand in the fire. A reason for doubting it would be that we know that Smith is engaged on research on fireproof lotions. We would require some such extraordinary reason before we could properly express doubt. Another reason would be that we doubt that fire burns. But circumstances would have to be very extraordinary indeed for us to doubt this. We would never normally say that one is presupposing that fire burns, though in very special circumstances we could come to say this.

(iii) We may doubt that a person is really in pain when he groans, if we have special reasons for doing so, e.g. the man is a confirmed practical joker. This case is different from the preceding one however in that it would never be proper to doubt that normally people do not disguise their emotions. That is, while in both cases it may in special circumstances be proper to doubt that the generalisation applies, in the second case it is not proper to doubt the generalisation itself. Similarly, it is never proper to doubt that things are in general what they seem to be, though it can in special circumstances be doubted that any particular thing is what it seems to be. (Both of these general statements may be what we find when we look for the ' presuppositions of language '.) In these cases, if we try to imagine the normal situation as being different, the language within which such a situation can be picked out seems to break up. How would it be if people were normally insincere in their expression of their feelings? If there were conventions governing the insincerity—e.g. if it were done to grin when in pain—the insincerity would become transparent and to all intents and purposes disappear. If there were no conventions governing insincerity, we would lose the possibility of expression. One could not learn the correct expression, and therefore could not put on an incorrect one. There would be no rules governing our demeanour and hence no deliberate breaking of rules. Expression of feeling and all talk of expression of feeling would be meaningless; which is the same as to say, there could be neither of these.

It may look here as if I am saying that certain things about the world must be as they are. But I do not want to deny that perhaps a world could be described in which there was no uniformity of reactions to such things as the touch of fire, or the approach of death, on the part of creatures who had what we would be inclined to call a language. All I want to say is that a doubt that their fellows are generally sincere in their expressions of feeling would not be possible to these creatures because

it would be inexpressible ; and that where such a doubt seems to be expressible, it is not a proper doubt.

<p style="text-align:center">*　　　*　　　*</p>

My conclusions are as follows :

(I) That an argument or statement of grounds is not deductively sufficient does not imply that there is any presupposition being made, or that there is any doubt.

Reasons are required for us to doubt that a case is normal, even where no attempt has been made to establish the fact that a case is normal. Where there are no such reasons, there can be no doubt ; where there is no doubt, there is no presupposition. " Circumstances are in this case normal " is not in general a presupposition of our judgments.

(II) A further doubt, that there is something wrong with our concept of the normal, is in some cases likely, in others hardly feasible. But in certain cases such a doubt is impossible because it destroys the context in which such a doubt would be expressed. It would therefore be incorrect to say that the correctness of our concept of the normal is a presupposition of all our judgments in such a context, or of our language.

What I have been saying applies to kinds of language other than those I have mentioned : for example, it has, I think, importance with regard to the problem of formulating general principles of morals which cover all possible cases.

University College of Wales, Aberystwyth.

BELIEVING THE MEANINGLESS : A REPLY TO MR. MELLOR

By AVRUM STROLL

IN ANALYSIS for October, 1953, I suggested a number of arguments in favour of saying that sentences of the form " X believes that p," when " p " is meaningless, are either (a) non-significant or (b) false. I rejected the possibility that such sentences could be true on the ground that if " p " failed to express a proposition there was nothing X could believe. To say, therefore, " X believes that p " either is a meaningless remark or is a mistaken one. Further, I indicated the arguments

supporting the latter alternative are stronger than those supporting the former, and for that reason the interpretation of such sentences as false is to be preferred.

Mr. W. W. Mellor in ANALYSIS for December 1954, takes exception to my statement that belief sentences of this kind cannot be true. He feels there is a use of " believe " in common speech such that it is true to say of someone " he believes that p," even though " p " may be non-significant. His view, if I understand him, is that " believes " refers to a " state of mind " in some of its uses, and accordingly that a sentence like " Santayana believed that the ideas have an independent subsistence ", may be a true description of Santayana's state of mind, even though " the ideas have an independent subsistence " is itself not meaningful. Mr. Mellor's reasons for holding this position seem to me to be mistaken and in this brief note I should like to say why.

I pointed out in my original paper that the problem of analyzing belief sentences which contain meaningless subsentences arises only for those philosophers who, for one reason or another, maintain that certain sentences are meaningless. Now it may be asked in opposition to such philosophers (Carnap, whom I quoted, is one) why common men ever bother to utter sentences which are meaningless. I suggested then that there are a number of answers of varying degrees of complexity to this question, but I did not wish to discuss the point at the time. But now some brief comments seem relevant. Besides the danger that one may commit the monistic fallacy in regarding every declarative sentence as significant, there are other reasons for thinking Carnap is right. For example, because of early training in (say) religion, or because they have been given courses in philosophy or literature which contained sentences they did not understand but which they could repeat, ordinary people may occasionally utter remarks like " I believe that p ", where " p " is meaningless. And it is possible, as in the cases of Heidegger and Hegel, that by a hypnotic extension of ordinary language, some philosophers may also on occasion talk gibberish, to which they prefix the words " I believe that ". Now Mr. Mellor's main argument in his paper seems to be that belief sentences containing meaningless subsentences cannot be either false or meaningless because they frequently have a use in everyday language. As he puts it :

> ' It seems to me that there are very grave objections to saying that such a sentence as " Santayana believed that the ideas have an independent subsistence " is either false or

meaningless. It cannot be meaningless, for it has a use in everyday language : it is equally odd to say that it is false. Would Mr. Stroll fail a candidate who wrote this sentence in an examination on Santayana's philosophy? We should all normally say that the sentence in question stated an important fact about Santayana. It tells us the kind of things Santayana tended to say and to write. (I shall call this "telling us something about Santayana's state of mind", though the expression is perhaps not a very happy one.)'

But this argument is unsatisfactory for a number of reasons. For one thing, it is not clear what "having a use in everyday language", means in this context. If "having a use" means only that certain sentences occur in everyday language in the sense that people sometimes repeat them, this is compatible with the possibility that such sentences may be meaningless for the reasons I mentioned previously. If, on the other hand, "having a use" means not only that certain sentences are repeated by people, but also that they are uttered by them for a certain purpose, then Mr. Mellor's claim that these sentences cannot be *cognitively* meaningless is much too strong, since nonsense may in this sense of "use" have a use in the everyday employment of language. For example, we may use nonsense sentences to entertain, excite, puzzle, or even to titillate someone aesthetically, as when a poet says "I believe that bits of green duration flutter in the room".

The second part of Mr. Mellor's argument is not much more defensible. When a student writes on an examination "Santayana believed that the ideas have an independent subsistence" a teacher would not ordinarily fail him, because generally he would be interpreted to mean (if he were more accurate) the same as "Santayana said that the ideas have an independent subsistence". Now this latter sentence which is not a belief sentence is indeed true. One must, however, carefully distinguish it from "Santayana believed that the ideas have an independent subsistence", which Mr. Mellor apparently fails to do, since he says : "It tells us the kind of things Santayana tended to *say* and *write*" (my italics). From the fact that Santayana tended to say and write certain sentences one cannot conclude, as I have tried to point out, that he was believing anything when he said or wrote them, even though he may have thought he was—assuming, of course, that such sentences are indeed meaningless.

There is, to be sure, a sense of "belief" in ordinary language

where one can say truly "A believes that p ", even where A is not now apprehending a proposition. Thus, it is true to say of Russell " he believes that $2+2 = 4$ ", even though he may not at the moment be aware of or be thinking about the proposition that $2+2 = 4$. In this sense, as is well known, " belief " is a dispositional term. But from this, we should not infer, as Mr. Mellor apparently would have us do, that one can be in the state of mind called " believing " when one writes or utters a sentence which expresses no proposition at all ; for as I originally argued, if there is nothing which is expressed by such a sentence, there is nothing which can be believed, even in a dispositional sense of " believe ".

If one wishes, finally, to rest his case on an appeal to ordinary language, as Mr. Mellor seems to, I say that ordinary men would be puzzled by his proposal that belief sentences of the sort we are considering can be true. If one is told " Heidegger believes that The Nothing Nots ", it seems natural to ask : " What is it that Heidegger believes? " Now it may be possible to answer this question, i.e., to translate the words " The Nothing Nots " into a significant sentence, and thus to explain what it is that Heidegger believes. But if this could not be done, and if in answer to the question " What is it that Heidegger believes? " the ordinary man then were told "Well, he believes but he really believes nothing at all ", he would be genuinely puzzled. Mr. Mellor is proposing an analysis which has the consequence that one may believe and yet believe nothing, and this in my opinion is highly implausible.

University of British Columbia.

is
of
ot
n
a
t.
e
ee
e
y
l
e